Genomic Prediction Models for Count Data

Osval A. Montesinos-López, Abelardo Montesinos-López, Paulino Pérez-Rodríguez, Kent Eskridge, Xinyao He, Philomin Juliana, Pawan Singh, and José Crossa

Whole genome prediction models are useful tools for breeders when selecting candidate individuals early in life for rapid genetic gains. However, most prediction models developed so far assume that the response variable is continuous and that its empirical distribution can be approximated by a Gaussian model. A few models have been developed for ordered categorical phenotypes, but there is a lack of genomic prediction models for count data. There are well-established regression models for count data that cannot be used for genomic-enabled prediction because they were developed for a large sample size (n) and a small number of parameters (p); however, the rule in genomic-enabled prediction is that p is much larger than the sample size n. Here we propose a Bayesian mixed negative binomial (BMNB) regression model for counts, and we present the conditional distributions necessary to efficiently implement a Gibbs sampler. The proposed Bayesian inference can be implemented routinely. We evaluated the proposed BMNB model together with a Poisson model, a Normal model with untransformed response, and a Normal model with transformed response using a logarithm, and applied them to two real wheat datasets from the International Maize and Wheat Improvement Center. Based on the criteria used for assessing genomic prediction accuracy, results indicated that the BMNB model is a viable alternative for analyzing count data.

Key Words: Bayesian analysis; Gibbs sampler; Count data; Genomic prediction; Data augmentation.

1. INTRODUCTION

Of all the computationally intensive methods for fitting complex multilevel models, the Gibbs sampler is most popular. Its popularity is due to its simplicity and its ability to effec-

Osval A. Montesinos-López is a Biometrician at the Biometrics and Statistics Unit of the International Maize and Wheat Improvement Center (CIMMYT), México, DF, México. Abelardo Montesinos-López is a PhD Student in the Departamento de Estadística, Centro de Investigación en Matemáticas (CIMAT), Guanajuato, 36240 Guanajuato, México. Paulino Pérez-Rodríguez is a Professor of Statistics at Colegio de Postgraduados, CP 56230 Montecillos, Edo. de México, México. Kent Eskridge is a Professor of Statistics, Deparment of Statistics at the University of Nebraska, Lincoln, NE 68583-0963, USA. Xinyao He, Pawan Singh are Pathologists at the Global Wheat Breeding Program of CIMMYT, Apdo. Postal 6-641, 06600 México, DF, México. Philomin Juliana is a PhD Student at the Plant Breeding & Genetics, Cornell University, 240 Emerson Hall, Cornell, Ithaca, NY, USA. José Crossa (✉) is a Biometrician at the Biometrics and Statistics Unit of CIMMYT, Texcoco, México
(E-mail: *j.crossa@cgiar.org*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 20, Number 4, Pages 533–554
DOI: 10.1007/s13253-015-0223-4

tively generate samples from a high-dimensional probability distribution (Park and van Dyk 2009). Despite these two advantages, we know of no efficient, closed-form Gibbs sampler for count data that is available for performing Poisson and negative binomial regression analyses. The Gibbs sampler proposed by Albert and Chib (1993) is built on a data augmentation approach and is one of the most widely used samplers for Bayesian probit regression. Recently, an analogous Gibbs sampler for Bayesian logistic regression was introduced by Polson et al. (2013). Their method differs from that of Albert and Chib (1993) in that they used Pólya–Gamma random variables instead of truncated normal random variables. Also, Polson et al. (2013) point out that their method can be applied to specific likelihoods related to the logistic function where it is possible to augment the joint density with auxiliary variables following a Pólya–Gamma distribution; this leads to a closed-form Gibbs sampler for binary and over-dispersed counts. However, Polson et al. (2013) focus most of their paper on binary outcomes. For this reason, in this paper, we provide a derivation of the closed-form Gibbs sampler for implementing a Bayesian mixed negative binomial (BMNB) regression model for counts applied in genomic selection.

Predicting yet-to-be observed phenotypes or unobserved genetic values for complex traits and inferring the underlying genetic architecture utilizing genomic data are interesting and fast developing areas in the context of plant and animal breeding, and even in human diseases (Goddard and Hayes 2009; de los Campos et al. 2010, 2013a; Riedelsheimer et al. 2012; Zhang et al. 2014). Rapid genetic progress requires that such predictions are accurate and can be produced early in life. For these reasons, the use of whole genome prediction models continues to increase.

In genomic prediction, all markers are simultaneously included in the model used for prediction. Real data analysis and simulation studies promote the use of this methodology for increasing genetic progress in less time. For continuous phenotypes, models have been developed to regress phenotypes on all available markers using a linear model (Zhang et al. 2014; de los Campos et al. 2013b). However, in plant breeding, the response variable in many traits is a count ($y = 0, 1, 2, \ldots$), for example, panicle number per plant, seed number per panicle, weed count per plot, number of infected spikelets per spike, etc. Statistical models used to analyze continuously distributed traits are not always optimal for analyzing categorical responses such as counts that are discrete, non-negative, integer-valued, and typically have right-skewed distributions. In classic probability theory, Poisson regression and negative binomial (NB) regression are often used to deal with count data. However, NB regression is preferred when the variance of the counts is larger than the mean, because the Poisson regression assumes that, conditional on any fixed values of the explanatory variables, the response mean and variance are equal. These models are different from an ordinary linear regression model. First, they do not assume that counts follow a normal distribution. Second, rather than modeling y as a linear function of the regression coefficients, they model a function of the response mean as a linear function of the coefficients. Regression models for counts are usually nonlinear and have to take into consideration the specific properties of counts, including discreteness and non-negativity, and are often characterized by overdispersion (variance greater than the mean).

Despite the special characteristics of discrete response data, it is still common practice, in the context of genomic selection, to apply linear regression models to such data or trans-

formed data (Montesinos-López et al. 2015). In genomic prediction, Kizilkaya et al. (2014) studied the reduction in model prediction accuracy for ordinal categorical traits relative to continuous traits. For smaller counts, data analysts are often advised to use logarithmic or square root transformations. There is mounting evidence that transformations do more harm than good for the models required by the vast majority of contemporary plant and soil science researchers (Stroup 2015). In this paper, we propose a NB regression model for counts using a data augmentation approach. We build on the fact that the gamma distribution is the conjugate prior of the NB parameter r, and that the NB random variable can be generated under a compound Poisson representation, which produces an efficient Gibbs sampling.

The article is organized as follows. In Sect. 2, we present the two datasets (Sect. 2.1) and the various models used (Sects. 2.2–2.5). The NB distribution is presented in Sect. 2.2, and the models applied to the two datasets are described in Sect. 2.3. Section 2.4 gives the prior distributions. Section 2.5 provides the full conditional distributions of the proposed models; details of model implementation are given in Sect. 2.6. In Sect. 2.7, the different criteria for assessing prediction accuracy are described. In Sect. 3, we illustrate the various methods with two real datasets and compare the proposed NB model with the Poisson and normal models. We discuss the results in Sect. 4 and finalize the study with the conclusions in Sect. 5.

2. MATERIALS AND METHODS

2.1. PHENOTYPE AND GENOTYPE DATA

The data used in this study are from the Global Wheat Program of the International Maize and Wheat Improvement Center (CIMMYT) and comprise the 46th (C46) and 47th (C47) International Bread Wheat Screening Nurseries (IBWSN) that were distributed worldwide in 2011 and 2012. The 297 lines from nursery C46 and the 425 lines from nursery C47 were evaluated for Fusarium Head Blight (FHB) resistance at El Batan Experiment Station, located at CIMMYT Headquarters near Texcoco (state of México, México). Ten spikes from each wheat line in the nurseries were tagged at anthesis using red sticky tape in the morning, followed by spray inoculation in the afternoon. The number of infected spikelets per spike was counted on each of the 10 tagged spikes; these numbers constitute the FHB count data.

Genotypes of the C46 and C47 lines were obtained with 45,000 Genotyping-By-Sequencing (GBS) markers following the protocol of Poland et al. (2012) where the absence or presence of marker genotypes of the wheat lines is represented by 0 and 1, respectively. We kept 13,913 and 13,120 GBS for C46 and C47 nurseries, respectively, that had $<50\,\%$ missing data; after deleting monomorphic markers with minor allele frequency (MAF) of $\leq$ 0.05, we ended up with a total of 11,218 and 11,510 GBS for the lines in the C46 and C47 nurseries, respectively. The remaining missing markers were imputed using the multivariate normal expectation maximization (EM) algorithm described in Poland et al. (2012).

2.2. NEGATIVE BINOMIAL DISTRIBUTION

Given that the NB distribution can arise in different ways, next we present its Gamma-Poisson representation. Let $Y|\mu \sim Pois\,(\mu)$ and $\mu|r, \pi \sim G\left(r, \frac{\pi}{1-\pi}\right)$, where $Pois\,(\mu)$ is

the Poisson distribution with mean and variance μ, and $G\left(r, \frac{\pi}{1-\pi}\right)$ is the distribution of a gamma random variable with shape parameter r and scale $\pi/(1-\pi)$, with $\pi \in (0, 1)$. It can be shown that the marginal distribution of Y has a probability mass function

$$\Pr(Y = y) = \frac{\Gamma(r+y)}{y!\Gamma(r)}(1-\pi)^r \pi^y, \text{ for } y = 0, 1, 2, \ldots \tag{2.1}$$

where $\Gamma(\cdot)$ denotes the gamma function. The resulting probability mass function in (2.1) corresponds to the *NB* distribution with parameters r and π, which from here on will be denoted as $NB(r, \pi)$. Therefore, the NB distribution is also known as the Gamma-Poisson distribution. The mean of this distribution $E(Y) = \mu = \frac{r\pi}{(1-\pi)}$ is smaller than variance $Var(Y) = \frac{r\pi}{(1-\pi)^2} = \mu + \frac{\mu^2}{r}$ with the variance-to-mean ratio denoted as $(1-\pi)^{-1}$ and the overdispersion level as r^{-1}. In terms of the mean parameter μ, we will use an alternative notation $\text{NB}(\mu, r)$. The NB distribution can also be generated using a Poisson representation (Quenouille 1949) as $Y = \sum_{l=1}^{L} u_l$, where $u_l \sim Log(\pi)$ and is independent of $L \sim Pois(-r\log(1-\pi))$, where Log and $Pois$ denote logarithmic and Poisson distributions, respectively. The distribution in (2.1) is also valid for any positive real value r.

2.3. MODELS FOR THE DATA

Except where otherwise noted, we use $i = 1, \ldots, n$ to index n lines, $j = 1, 2, \ldots, m_i$ to index m_i spikes for the ith line, and $k = 1, 2, \ldots, p$ to index p markers. We use y_{ij} to represent the number of infected spikelets for the jth spike of the ith line, and x_{ik} to represent the genotype of the ith line at the kth single-nucleotide polymorphism (SNP) marker. (For a given marker, the genotype for the ith line is coded as the number of copies of a designated marker-specific allele carried by the ith line). Each of the data models we consider involves the linear predictor

$$\eta_i = \beta_0 + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i \tag{2.2}$$

where β_0 is the intercept parameter, $\boldsymbol{x}_i^{\mathrm{T}} = \left[x_{i1}, \ldots x_{ip}\right]$ is the marker genotype information for the ith line, $\boldsymbol{\beta}^{\mathrm{T}} = \left[\beta_1, \ldots, \beta_p,\right]$ is a vector of fixed allele substitution effects, and u_i is a random effect for the ith line that represents the genetic value of line i not captured by genotypes at the p markers. We propose the following four models for analyzing the wheat dataset described in Sect. 2.1.

Model NB: $y_{i1}, \ldots, y_{im_i} | \eta_i \overset{i.i.d}{\sim} \text{NB}(\mu_i, r)$, with r being the overdispersion parameter, $\mu_i = \exp(\eta_i)$.

Model Pois: Similar to **Model NB**, except that $y_{i1}, \ldots, y_{im_i} | \eta_i \overset{i.i.d}{\sim} \text{Pois}(\mu_i)$.

Model Normal: Similar to **Model NB**, except that $y_{i1}, \ldots, y_{im_i} | \eta_i \overset{i.i.d}{\sim} N(\eta_i, \sigma_e^2)$ with identity link function.

Model Log-normal: Similar to **Model NB**, except that $\log\left(y_{i1}+1\right), \ldots, \log\left(y_{im_i}+1\right) | \eta_i \sim N(\eta_i, \sigma_e^2)$ with identity link function.

Note that in genomic-enabled prediction, **Model Normal** without random effects is called Bayesian Ridge Regression (BRR) when the sparseness is included in the model

by specifying a $\boldsymbol{\beta} \sim N_p\left(\mathbf{0}, \boldsymbol{I}\sigma_\beta^2\right)$ prior for the parameter $\boldsymbol{\beta}$. Also, **Model Normal** without the $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ term is called the Genomic Best Linear Unbiased Predictor (GBLUP) model (de los Campos et al., 2013b) and $\boldsymbol{u}$ denotes the additive genetic values of lines. Under the Bayesian GBLUP, the prior density of the genetic values $\boldsymbol{u} = (u_1, \ldots, u_n)^{\mathrm{T}} \sim N\left(\mathbf{0},\ \boldsymbol{G}\sigma_u^2\right)$ is a conjugate multivariate normal $N\left(\mathbf{0},\ \boldsymbol{G}\sigma_u^2\right)$, where the matrix $\boldsymbol{G}$ is estimated from marker data $\boldsymbol{X}$ (for $k = 1, 2, \ldots, p$ markers) as $\boldsymbol{G} = \frac{\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}}{p}$ (VanRaden 2007, 2008) and called the genomic relationship matrix (GRM). In the next section, we describe the implementation of Bayesian mixed negative binomial (BMNB) regression.

2.4. PRIOR DISTRIBUTIONS

Considering **Model NB**, note that conditionally on u_i, the probability that the random variable Y_{ij} takes the value y_{ij} can be expressed as

$$\begin{aligned}\Pr\left(Y_{ij} = y_{ij}\,|u_i\right) &= \frac{\Gamma\left(y_{ij}+r\right)}{y_{ij}!\Gamma\left(r\right)}\left(1-\frac{\mu_i}{r+\mu_i}\right)^r\left(\frac{\mu_i}{r+\mu_i}\right)^{y_{ij}} \\ &= \frac{\Gamma\left(y_{ij}+r\right)}{y_{ij}!\Gamma\left(r\right)}\frac{\left[\exp\left(\eta_{ij}^*\right)\right]^{y_{ij}}}{\left[1+\exp\left(\eta_{ij}^*\right)\right]^{y_{ij}+r}}\ (\text{for } y_{ij} = 0, 1, 2, \ldots) \qquad (2.3)\end{aligned}$$

since $\pi_{ij} = \frac{\mu_i}{r+\mu_i} = \frac{\exp(\eta_i^*)}{1+\exp(\eta_i^*)}$, where $\eta_i^* = \beta_0^* + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i$, with $\beta_0^* = \beta_0 - \log\left(r\right)$. Then

$$\begin{aligned}\Pr\left(Y_{ij} = y_{ij}\,|u_i\right) &= \frac{\Gamma\left(y_{ij}+r\right)}{y_{ij}!\Gamma\left(r\right)}2^{-y_{ij}-r}\exp\left(\frac{y_{ij}-r}{2}\eta_i^*\right)\int_0^\infty \\ &\exp\left[-\frac{\omega_{ij}\left(\eta_i^*\right)^2}{2}\right]P\left(\omega_{ij};\, y_{ij}+r, 0\right)d\omega_{ij}\end{aligned}$$

which was obtained using the equality given by Scott and Pillow (2013): $\frac{\left(e^\psi\right)^a}{\left(1+e^\psi\right)^b} = 2^{-b}e^{\kappa\psi}\int_0^\infty e^{-\frac{\omega\psi^2}{2}}P\left(\omega; b, 0\right)d\omega$, where $\kappa = a - b/2$ and $P\left(.; b, 0\right)$ is the density of $PG\left(b, c = 0\right)$, the Pólya–Gamma distribution with parameters b and $c = 0$ (see Definition 1 in Polson et al. 2013).

From here, conditional on $\omega_{ij} \sim PG\left(y_{ij}+r, 0\right)$,

$$\Pr\left(Y_{ij} = y_{ij}\,|u_i, \omega_{ij}\right) = \frac{\Gamma\left(y_{ij}+r\right)}{y_{ij}!\Gamma\left(r\right)}2^{-y_{ij}-r}\exp\left(\frac{y_{ij}-r}{2}\eta_i^*\right)\exp\left[-\frac{\omega_{ijt}\left(\eta_i^*\right)^2}{2}\right]. \qquad (2.4)$$

To complete the Bayesian specification, here we provide the prior distributions, $f\left(\boldsymbol{\theta}\right)$, for all the unknown model parameters β_0^*, $\boldsymbol{\beta}$, σ_β^2, $\boldsymbol{u}$, σ_u^2, and r. We assume prior independence between the parameters, that is,

$$f\left(\boldsymbol{\theta}\right) = f\left(\beta_0^*\right)f\left(\boldsymbol{\beta}\right)f\left(\sigma_\beta^2\right)f\left(\boldsymbol{u}\right)f\left(\sigma_u^2\right)f\left(r\right)$$

The prior specification in terms of β_0^* instead of β_0 is for convenience, as we shall see in what follows. Since we have no prior information, we assign conditionally conjugate but weakly informative prior distributions to the parameters. Also, to guarantee proper posteriors, we adopt proper priors with known hyper-parameters whose values we specify in Sect. 2.6. We assume that $\beta_0^* \sim N\left(\beta_f, \sigma_0^2\right)$, $\boldsymbol{\beta} \sim N_p\left(\boldsymbol{\beta}_v, \boldsymbol{\Sigma}_v\sigma_\beta^2\right)$, $\sigma_\beta^2 \sim IG\left(a_\beta, b_\beta\right)$, where $IG\left(a_\beta, b_\beta\right)$ denote the inverse gamma distribution with shape a_β and scale b_β parameters, $\boldsymbol{u} \sim N_n\left(\mathbf{0}, \boldsymbol{G}\sigma_u^2\right), \sigma_u^2 \sim IG\left(a_u, b_u\right)$, and $r \sim G\left(a_0, 1/b_0\right)$. Next we combine (2.4) using all data with priors to get the full conditional distribution for the parameters β_0^*, $\boldsymbol{\beta}$, σ_β^2, $\boldsymbol{u}$, σ_u^2, and r.

2.5. Full Conditional Distributions

From (2.4) and the prior specification given in the previous section, the full conditional for β_0^* is

$$f\left(\beta_0^*|.\right) \propto \exp\left(-\frac{1}{2\tilde{\sigma}_0^2}\left(\beta_0^* - \tilde{\beta}_0\right)^2\right) \propto N\left(\tilde{\beta}_0, \tilde{\sigma}_0^2\right) \tag{2.5}$$

where $\tilde{\sigma}_0^2 = \left(\sigma_0^{-2} + \mathbf{1}_m^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\right)^{-1}$, $\tilde{\beta}_0 = \tilde{\sigma}_0^2\left(\sigma_0^{-2}\beta_f - \mathbf{1}_m^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_m^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Z}\boldsymbol{u} + \mathbf{1}_m^{\mathrm{T}}\boldsymbol{\kappa}\right)$, $\boldsymbol{\kappa} = \left[\boldsymbol{\kappa}_1^{\mathrm{T}}, \ldots, \boldsymbol{\kappa}_n^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{\kappa}_i = \frac{1}{2}\left[y_{i1} - r, \ldots, y_{im_i} - r\right]^{\mathrm{T}}$, $\boldsymbol{X} = \left[\boldsymbol{X}_1^{\mathrm{T}}, \ldots, \boldsymbol{X}_n^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{X}_i = \left[\mathbf{1}_{m_i}^{\mathrm{T}} \otimes \boldsymbol{x}_i\right]^{\mathrm{T}}$, $\boldsymbol{D}_{\boldsymbol{\omega}} = \mathrm{diag}\left(\boldsymbol{D}_{\omega 1}, \ldots, \boldsymbol{D}_{\omega n}\right)$, $\boldsymbol{D}_{\omega i} = \mathrm{diag}\left(\omega_{i1}, \ldots, \omega_{im_i}\right)$, $\mathbf{1}_m = \left[\mathbf{1}_{m_1}^{\mathrm{T}}, \ldots, \mathbf{1}_{m_n}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{u} = [u_1, \ldots, u_n]^{\mathrm{T}}$, $\boldsymbol{Z} = \begin{bmatrix} \mathbf{1}_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{m_2} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{m_n} \end{bmatrix}$ and $\mathbf{1}_{m_i} = [1, \ldots, 1]^{\mathrm{T}}$ is a vector of dimension m_i with entries equal to 1. If we use a prior for $\beta_0^* \propto$ constant (improper uniform prior), then in $\tilde{\sigma}_0^2$ and $\tilde{\beta}_0$, we set the term σ_0^{-2} to zero (further details in Appendix 1).

The full conditional distribution of ω_{ij} is

$$f\left(\omega_{ij}|.\right) \sim PG\left(y_{ij} + r, \beta_0^* + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i\right) \tag{2.6}$$

The full conditional distribution for $\boldsymbol{\beta}$ is as follows:

$$f\left(\boldsymbol{\beta}|.\right) \propto \exp\left(-\frac{1}{2}\left[\left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_v\right)^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}_v^{-1}\left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_v\right)\right]\right) \propto N_p\left(\tilde{\boldsymbol{\beta}}_v, \tilde{\boldsymbol{\Sigma}}_v\right) \tag{2.7}$$

where $\tilde{\boldsymbol{\Sigma}}_v = \left(\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{X}\right)^{-1}$, $\tilde{\boldsymbol{\beta}}_v = \tilde{\boldsymbol{\Sigma}}_v\left(\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2}\boldsymbol{\beta}_v - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\beta_0^* - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\kappa}\right)$. Also, if here we use a prior for $\boldsymbol{\beta} \propto$ constant (improper uniform prior), then in $\tilde{\boldsymbol{\Sigma}}_v$ and $\tilde{\boldsymbol{\beta}}_v$, we set the term $\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2}$ to zero.

The full conditional distribution for σ_β^2 is

$$f\left(\sigma_\beta^2|.\right) \sim IG\left(a_\beta + \left(\boldsymbol{\beta} - \boldsymbol{\beta}_v\right)^{\mathrm{T}} \sum\nolimits_{\boldsymbol{v}}^{-1} \left(\boldsymbol{\beta} - \boldsymbol{\beta}_{\boldsymbol{v}}\right)/2, b_\beta + p/2\right) \tag{2.8}$$

The full conditional distribution for $\boldsymbol{u}$ is

$$f\left(\boldsymbol{u}|.\right) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{u} - \tilde{\boldsymbol{u}}\right)^{\mathrm{T}} \boldsymbol{F}^{-1}\left(\boldsymbol{u} - \tilde{\boldsymbol{u}}\right)\right\} \propto \quad N_n\left(\tilde{\boldsymbol{u}},\ \boldsymbol{F}\right) \tag{2.9}$$

where $\boldsymbol{F} = \left(\sigma_u^{-2}\boldsymbol{G}^{-1} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_\omega\boldsymbol{Z}\right)^{-1}$ and $\tilde{\boldsymbol{u}} = \boldsymbol{F}\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{k} - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_\omega\mathbf{1}_m\beta_0^* - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_\omega\boldsymbol{X}\boldsymbol{\beta}\right)$.

The conditional distribution of σ_u^2 is

$$f\left(\sigma_u^2|.\right) \sim IG\left(a_u + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{u}/2, b_u + n/2\right) \tag{2.10}$$

The full conditional for r is not known, and the development of a Metropolis-Hasting step for this implies evaluating the density of a Pólya–Gamma distribution. This evaluation cannot be done directly because an alternating infinite series is involved. Therefore, after obtaining a sample of β_0^*, $\boldsymbol{\beta}$, σ_β^2, and σ_u^2 given r, we can adopt the strategy of Zhou et al. (2012) to obtain a sample of the full conditional of r by alternating

$$f\left(r|.\right) \sim G\left(a_0 + \sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{m_i} L_{ij},\ \frac{1}{b_0 - \sum_{i=1}^{n}\sum_{j=1}^{m_i}\log\left(1 - \pi_{ij}\right)}\right) \tag{2.11}$$

$$f\left(L_{ij}|.\right) \overset{iid}{\sim} CRT\left(y_{ij}, r\right) \tag{2.12}$$

where $CRT\left(y_{ij}, r\right)$ denote the Chinese restaurant table (CRT) count random variable and can be generated as $L_{ij} = \Sigma_{l=1}^{y_{ij}} d_l$, where $d_l \overset{iid}{\sim} Bernoulli\left(\frac{r}{l-1+r}\right)$ (Zhou and Carin 2015). Further details on the full conditional distributions for these parameters are given in Appendix 1.

In summary, samples of the joint posterior distribution of all the parameters involved can be obtained by sampling repeatedly from the following loop:

1. Sample ω_{ij} values from the Pólya-Gamma distribution in (2.6).
2. Sample β_0^* from the normal distribution in (2.5).
3. Sample $\boldsymbol{\beta}$ from the normal distribution in (2.7).
4. Sample the variance effect (σ_β^2) from the inverse gamma distribution in (2.8).
5. Sample $\boldsymbol{u}$ from the normal distribution in (2.9).
6. Sample the variance effect (σ_u^2) from the inverse gamma distribution in (2.10).
7. Sample the parameter (r) from the gamma distribution in (2.11) and (2.12).

Return to step 1 or terminate if chain length is adequate to meet convergence diagnostics.

The Gibbs sampler proposed above without random effects $(\boldsymbol{u})$ and assuming a normal distribution $N_p(\mathbf{0}, \boldsymbol{I}_p\sigma_\beta^2)$ as prior of the parameter $\boldsymbol{\beta}$ produces the following conditional posterior expectation

$$E\left(\boldsymbol{\beta}|\boldsymbol{y}, r, \boldsymbol{\omega}\right) = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{X} + \boldsymbol{I}_p\sigma_{\beta}^{-2}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{\kappa} - \boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\beta_0^*\right)$$

which is analogous to the BRR with normal phenotype but with pseudo response $\boldsymbol{y}^* = \boldsymbol{\kappa} - \boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\beta_0^*$. This can be viewed as Bayesian ridge regression for counts (count BRR). Implementation of the count BRR model is straightforward using the Gibbs sampler proposed above but ignoring steps 5 and 6. On the other hand, ignoring the term $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ ($\boldsymbol{X}\boldsymbol{\beta}$) in the linear predictor of (2.2) and defining $u_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ gives the GBLUP count, since the posterior expectation of the additive genetic values is equal to $E\left(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\omega}\right) = \left(\sigma_u^{-2}\boldsymbol{G}^{-1} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_{\omega}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\mathrm{T}}\left(\boldsymbol{k} - \boldsymbol{D}_{\omega}\mathbf{1}_m\beta_0^*\right)$. The GBLUP count with the proposed Gibbs sampler can also be implemented by ignoring steps 3 and 4 above. Both models (BRR count and GBLUP count) are expected to produce the same results, since they are different parameterizations of the same model (de los Campos et al. 2010).

Since $\lim_{r\to\infty} NB\left(r, \frac{\mu}{\mu+r}\right) = \text{Pois}\left(\mu\right)$, **Model Pois** was implemented with the above method fixing r to a large value depending on the mean count. We used $r = 1000$, which is a good choice when the mean count is less than 100. Also, **Model Normal** and **Model log-normal** were implemented under a Bayesian framework following Kärkkäinen and Sillanpää (2012).

2.6. MODEL IMPLEMENTATION

The Gibbs sampler described above for the BMNB model (**Model NB**) was implemented in the R-software (R Core Team 2015). Implementation was done under a Bayesian approach using Markov Chain Monte Carlo (MCMC) through the Gibbs sampler algorithm, which samples sequentially from the full conditional distributions until it reaches a stationary process, converging to the joint posterior distribution (Gelfand and Smith 1990). To decrease the potential impact of MCMC errors on prediction accuracy, we performed a total of 60,000 iterations with a burn-in of 30,000, so that 30,000 samples were used for inference. We did not apply thinning of the chains following the suggestions of Geyer (1992), MacEachern and Berliner (1994), and Link and Eaton (2012), who provide justification of the ban on subsampling MCMC output for approximating simple features of the target distribution (e.g., means, variances and percentiles) since thinning is neither necessary nor desirable, and unthinned chains are more precise.

We implemented the prior specification given in Sect. 2.4 with $\beta_0^* \sim N\left(\beta_f = 0, \sigma_0^2 = 10000\right)$, given σ_β^2 we take $\boldsymbol{\beta} \sim N_p\left(\boldsymbol{\beta}_v = \mathbf{0}_p^{\mathrm{T}}, \boldsymbol{I}_p\sigma_\beta^2\right)$, $\sigma_\beta^2 \sim IG\left(a_\beta = 0.01, b_\beta = 0.01\right)$, given σ_u^2 we take $\boldsymbol{u} \sim N_n\left(\mathbf{0}_n^{\mathrm{T}}, \boldsymbol{G}\sigma_u^2\right)$, $\boldsymbol{G}$ is the GRM, that is, the covariance matrix of the random effects, $\sigma_u^2 \sim IG\left(a_u = 0.01, b_u = 0.01\right)$ and $r \sim G\left(a_0 = 0.01, 1/(b_0 = 0.01)\right)$. All these hyper-parameters were chosen to lead weakly informative priors. The convergence of the MCMC chains was monitored using trace plots and autocorrelation functions. Also, when considering a sensitivity analysis on the use of the inverse gamma priors for the variance components, we found that the results are fairly robust under different choices of prior.

2.7. ASSESSING PREDICTION ACCURACY

We used cross-validation to estimate the prediction accuracy of the proposed models for count phenotypes. The dataset was divided 10 times into training and validation sets, with 90% of the dataset used for training and 10% for testing (since each line has 10 replications, we used 9 replicates for training and one replicate for testing). The training set was used to fit the model and the validation set was used to evaluate the prediction accuracy of the proposed models. Among the variety of methods for comparing the predictive posterior distribution to the observed data (generally termed "posterior predictive checks"), we used five criteria: the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002), the sum of the logged conditional predictive ordinate, also known as log-marginal pseudo-likelihood (LMPL) (Gelfand 1996), the Chi-Square statistic, χ^2_{cal}, (Gelman et al. 2004), the L criterion (L) (Laud and Ibrahim 1995), and the Pearson correlation (Corr). Models with small DIC, χ^2_{cal}, and L indicate better fitting, and a higher LMPL, and absolute values of Corr also indicate better fitting. The predicted observations for **Models NB and Pois**, $\hat{y}_{ij}$, were calculated with M collected Gibbs samples as $\hat{y}_{ij} = \frac{1}{M}\sum_{s=1}^{M}\exp\left[\beta_0^{*(s)} + \log(r^{(s)}) + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(s)} + u_i^{(s)}\right]$, where $\beta_0^{*(s)}$, $r^{(s)}$, $\boldsymbol{\beta}^{(s)}$ and $u_i^{(s)}$ are the values in the sample s of the intercept, overdispersion parameter, the beta regression coefficients, and the random effect for the line i, respectively. For the **Model Normal**, the $\hat{y}_{ij}$ were calculated as $\hat{y}_{ij} = \frac{1}{M}\sum_{s=1}^{M}\left[\beta_0^{(s)} + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(s)} + u_i^{(s)}\right]$, while for **Model Log-normal**, the $\hat{y}_{ij}$ were calculated as $\hat{y}_{ij} = \frac{1}{M}\sum_{s=1}^{M}\exp\left[\beta_0^{(s)} + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(s)} + u_i^{(s)} + \sigma_e^{2(s)}/2\right] - 1$, where $\sigma_e^{2(s)}$ is the value of the variance component of the error in sample s.

3. RESULTS

Figure 1a depicts the histogram of the number of infected spikelets per spike for the entire C46 dataset. The mean and variance of this dataset were 2.4721 and 5.8438, respectively, while the minimum and maximum numbers of spikelets on a spike were 0 and 13, respectively. Figure 1b shows that in 213 lines, the mean is smaller than the variance. However, this data set, C46, has 702 zero counts (23.63%), which always causes problems for fitting the Poisson and NB models. In dataset C47, the mean and variance were 3.2451 and 6.9594, respectively (Fig. 1c) and the minimum and maximum numbers of spikelets on a spike were 0 and 19, respectively. Also in dataset C47, 195 out of 425 lines had a mean that is smaller than the variance (Fig. 1d).

Table 1 shows a comparison of the four models used for the five proposed criteria using the full datasets (no random partitions were used). Concerning criterion DIC, for dataset C46, the best and second best models were **Model Log-normal** and **Model NB**, respectively, and for dataset C47, the best two models were **Model Log-normal** and **Model Pois**. In terms of χ^2_{cal}, the best model was **Model NB** in both datasets. With the L criteria, the best model was **Model Pois** in both data sets. For the LMPL criterion, the best model was **Model Log-normal** in both datasets. Finally, with the Pearson correlation (Corr), the best model for both datasets was **Model Log-normal**. We also got the average of the ranks of the five proposed criteria for each model and, in this situation, the rank of models for dataset C46

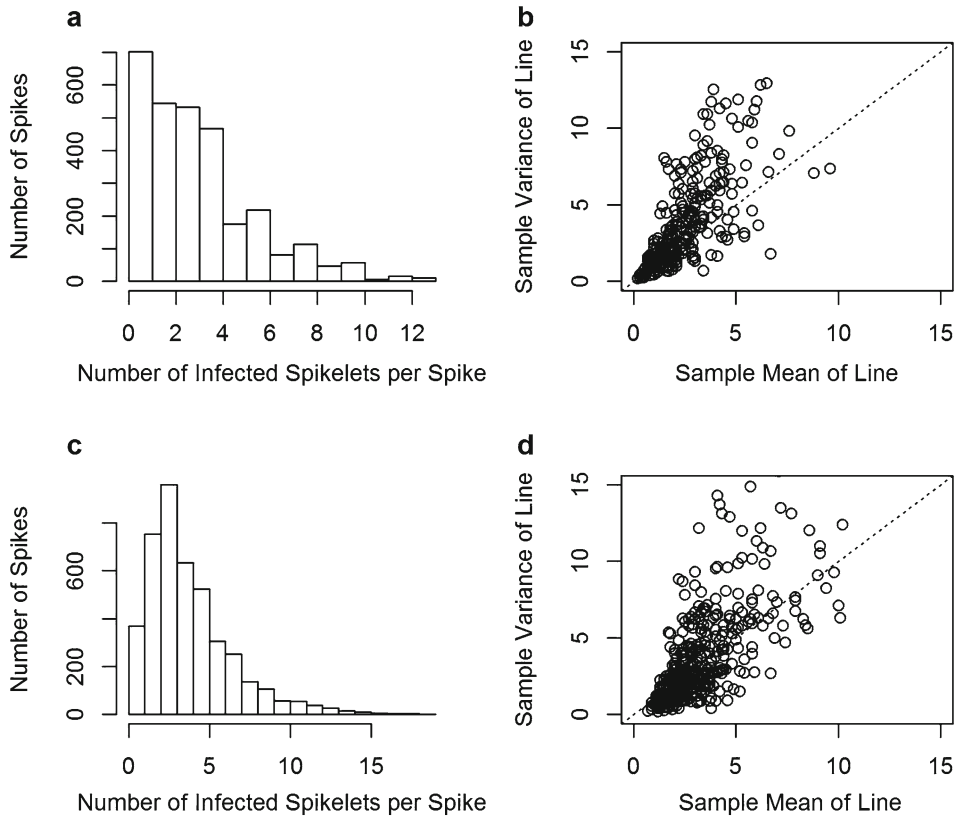

Figure 1. Histograms of observed counts of *Fusarium* resistance datasets C46 (**a**) and C47 (**c**). Scatterplots (**b**) for C46 and (**d**) for C47 were built with one point per line and the sample mean and sample variance for the ith line were computed from the observed counts $y_{i1}, \ldots, y_{im_i}$.

was 1 for **Model Log-normal**, 2 for **Model Pois**, and 3 for **Models NB** and **Normal**. For data set C47, the models ranked as follows: 1 for **Model Log-normal**, 2 for **Model Pois**, 3 for **Model Normal,** and 4 for **Model NB**.

Also for the four models, we show the histogram representation of the posterior distributions for scalar parameters for both data sets. In all plots in Figs. 2 (for dataset C46) and 3 (for dataset C47), it can be observed that the priors for each parameter in **Models NB, Pois, Normal**, and **Log-normal** are not informative.

In Table 2, we present the results of 10 cross-validations, where prediction accuracy was also assessed with the five proposed criteria. However, here these metrics were calculated using only the testing set (and not the whole dataset, as in Table 1). In Table 3, we present the ranking of the four models for each criterion. Since we are comparing four models, the values of the ranks range from 1 to 4, and the lower the values, the better the model. For ties, we assigned the average of the ranks that would have been assigned had there been no ties. From the ranking given in Table 3, we can see that there is no clear winner in terms of prediction accuracy. For example, in DIC and LMPL, the best model was **Model NB** for dataset C46, whereas for dataset C47, the best model was **Model Pois** in three criteria.

Table 1. Mean and standard deviation (in parentheses) of posterior distributions of the parameters β_0^*, r or σ_e^2, σ_u^2, and σ_β^2 for **Models NB, Pois, Normal**, and **Log-normal** using the full data for datasets C46 and C47.

	Parameter	**Model NB**	**Model Pois**	**Model Normal**	**Model Log-normal**
C46	β_0^*	−0.92 (0.08)	−6.26(0.04)	2.47(0.05)	0.72(0.02)
	r	4.87(0.45)	1000	3.76(0.10)	0.34(9E−3)
	σ_u^2	0.55(0.80)	0.29(0.15)	0.29(0.66)	0.11(0.15)
	σ_β^2	2E-3(2E-3)	4E−3(5E−3)	5E−3(4E−3)	4E−3(4E−3)
	DIC	11921.57[2]	12044.88[3]	12658.65[4]	10601.95[1]
	χ^2_{cal}	2208.89[1]	3580.08[4]	2427.72[3]	2280.57[2]
	L	8293.25[4]	6864.92[1]	7773.45[3]	7009.80[2]
	LMPL	−6444.10[4]	−6302.17[2]	−6337.94[3]	−5287.97[1]
	Corr	0.61[4]	0.62[3]	0.65[2]	0.71[1]
C47	β_0^*	−2.26(0.22)	−5.89(0.02)	3.25(0.05)	0.95(0.03)
	r	26.95(6.03)	1000	4.41(0.60)	0.31 (7E−3)
	σ_u^2	0.48(0.61)	0.06(0.04)	0.73(1.13)	0.32(0.55)
	σ_β^2	1E-3(1E-3)	3E−3(3E−3)	3E−3(3E−3)	3E−3(3E−3)
	DIC	18643.51[4]	16920.49[2]	18072.83[3]	15724.36[1]
	χ^2_{cal}	2588.80[1]	3845.58[4]	3041.74[3]	2922.00[2]
	L	11966.26[4]	10307.01[1]	11663.63[3]	11550.18[2]
	LMPL	-10140.1[4]	−8510.79[2]	−9119.42[3]	−7838.09[1]
	Corr	0.71[3]	0.71[3]	0.71[3]	0.73[1]

The [i] with $i = 1, 2, 3, 4$ denotes the ranking of the four models according to each of the five criteria: DIC, χ^2_{cal}, L, LMPL, and Corr. The lower the value of [i], the better the model. In **Models Normal** and **Log-normal**, the r parameter represents the σ_e^2.

The rank given by the χ^2_{cal} criterion is near the opposite of the rank given by criteria DIC and L; for χ^2_{cal}, the best model was **Model Log-normal** in both data sets. In terms of the L criterion, the best model was **Model Pois** for both datasets, while in terms of Corr, the best model was **Model Normal** for dataset C46, and all **Models** (**NB**, **Pois**, **Normal**, and **Log-normal**) for dataset C47. We also got the average of the ranks for each model of the five criteria, which is given in the last row in Table 3. In terms of the average ranking for dataset C46, the best model was **Model NB**, followed by **Model Normal**; **Model Pois** came in last. In dataset C47, the best model was **Model Pois**, the second best was **Model Normal**, and the last one was **Model Log-normal**.

Model Pois often gave better results than **Model NB**, even though **Model Pois** is a special case of **Model NB** with r fixed at 1000. It seems surprising that the posterior of r in these cases concentrates at values quite far from 1000. As a check to show that this behavior is possible, we performed a small simulation experiment with the following linear predictor $\eta_i = \beta_0 + u_i$ with $i = 1, \ldots, 40$ to index lines, $j = 1, 2, \ldots, 10$ to index spikes per line. In the first scenario, we simulated data according to **Model NB** with hyper-parameters equal to $\beta_0 = 0.66$, $r = 4.87$, and $\sigma_u^2 = 0.55$. In the second scenario, we simulated data according to **Model Pois** with hyper-parameters $\beta_0 = 0.66$, $r = 1000$, and $\sigma_u^2 = 0.55$. We expect that **Model NB** will work well and will be able to capture the true parameters when fitting **Model**

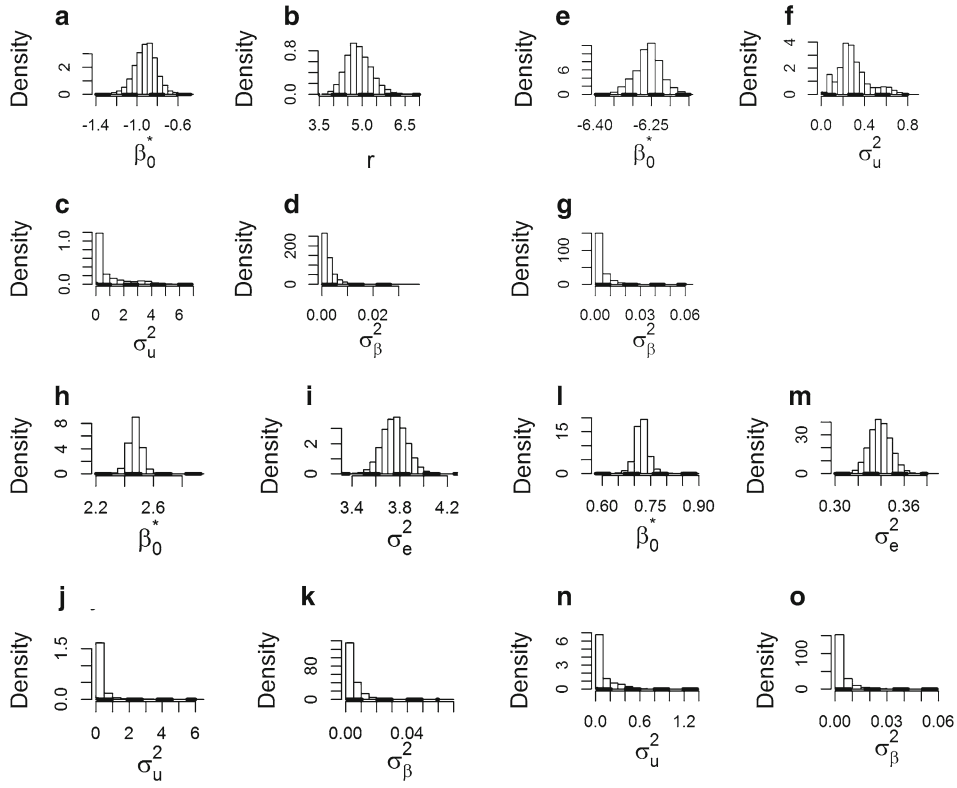

Figure 2. Histogram representation of posterior distributions of **Models NB (a–d), Pois (e–g), Normal (h–k) and Log-normal (l–o)** for dataset C46 for scalar parameters β_0^*, r or σ_e^2, σ_u^2 and σ_β^2 with priors superimposed as dashed lines at the bottom.

NB to data generated according to **Model NB**. We also expect that **Model Pois** will work well and will be able to capture the true parameters when fitted to data generated according to **Model Pois**. Table 4 gives the estimates of fitting both models (**Models NB** and **Pois**) to each simulated dataset; we also calculated the DIC, χ^2_{cal}, L, LMPL, and Corr for each fitted model to compare the models' performance.

Results shown in Table 4 are based in 50 replications and in each replication we computed 20,000 MCMC samples. Bayes estimates were computed with 10,000 samples since the first 10,000 were discarded as burn-in. Table 4 shows that the estimates obtained by fitting **Models NB** and **Pois** are close to the true values when the data were simulated under **Model NB**, with the exception of the parameter r, which was fixed under **Model Pois**. Upon comparing the performance of both models, we see that fitting **Model NB** to data generated according to **Model NB** gave better performance, since in three of the five criteria, this model is the best. On the other hand, for the data simulated according to **Model Pois**, we see that the estimates of β_0 and σ_u^2 are close to the true values when fitting both **Models NB** and **Pois**. However, the estimate of the parameter r was 61.94 when fitting **Model NB**, that is, a value far from the fixed value of $r = 1000$ that is assumed for **Model Pois**.

The reason that the estimated value of r is far from 1000 is because, for small counts, the r required to approximate the Poisson distribution with a negative binomial distribution is

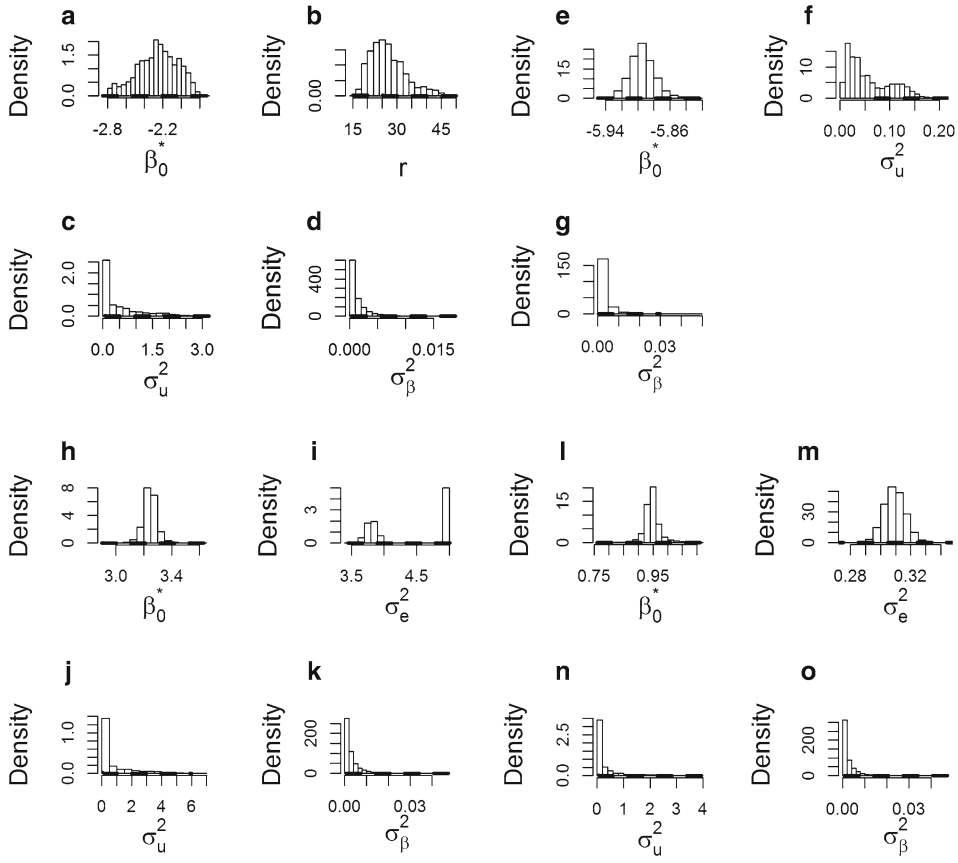

Figure 3. Histogram representation of posterior distributions of **Models NB (a-d), Pois (e-g), Normal (h-k), and Log-normal (l-o)** for dataset C47 for scalar parameters β_0^*, r *or* σ_e^2, σ_u^2, and σ_β^2 with priors superimposed as dashed lines at the bottom.

small. Comparing the performance of both models with the five criteria, we see that in DIC and χ^2_{cal}, the best is **Model NB**, while in L and LMPL, **Model Pois** was the best. However, **Model Pois** is preferred because it has one less parameter to be estimated. With this small simulation example, we observed that **Model NB** performed better when it was applied to data generated from **Model NB**. Likewise, the more parsimonious **Model Pois** performed adequately when applied to data generated according to **Model Pois**. For this reason, we conclude that the results (given above) produced by using real data are congruent.

4. DISCUSSION

4.1. THE POISSON AND NEGATIVE BINOMIAL MODELS THROUGH DATA AUGMENTATION

The proposed Bayesian regression models for count data take into account the nonlinear relationships between responses and consider the specific properties of counts, including discreteness, non-negativity, and overdispersion (variance greater than the mean). Although

Table 2. Mean, minimum and maximum values from the 10 cross-validation partitions for the five criteria (DIC, χ^2_{cal}, L, LMPL, and Corr) of each model (**Models NB, Pois, Normal** , and **Log-normal**) in each dataset, C46 and C47.

Model	Criterion	Datasets					
		C46			C47		
		Mean	Min	Max	Mean	Min	Max
Model NB	DIC	1243.18	1201.21	1279.57	1908.53	1715.36	2262.24
	χ^2_{cal}	334.76	282.54	401.27	365.26	314.27	406.38
	L	839.59	828.36	848.70	1208.28	1162.47	1358.77
	LMPL	−651.63	−679.21	−626.63	−1077.64	−1794.09	−943.46
	Corr	0.53	0.46	0.57	0.63	0.58	0.68
Model Pois	DIC	1306.81	1242.04	1473.22	1803.25	1739.62	1901.77
	χ^2_{cal}	500.19	421.78	555.48	498.43	412.92	615.82
	L	724.72	695.66	807.23	1080.84	1053.06	1121.79
	LMPL	−711.99	−986.05	−651.29	−916.87	−975.76	−880.21
	Corr	0.53	0.44	0.57	0.63	0.58	0.68
Model Normal	DIC	1270.26	1198.39	1318.33	1806.39	1760.73	1847.53
	χ^2_{cal}	182.42	136.39	218.59	468.27	369.39	620.089
	L	875.57	825.65	929.03	1403.43	1351.81	1453.88
	LMPL	−655.70	−678.62	−621.78	−932.90	−951.45	−920.16
	Corr	0.63	0.55	0.70	0.63	0.59	0.69
Model Log-normal	DIC	1367.84	1312.46	1433.32	1954.15	1865.66	2058.35
	χ^2_{cal}	135.11	122.15	144.71	304.77	282.58	323.67
	L	908.45	870.56	944.15	1413.77	1338.05	1463.74
	LMPL	−686.12	−720.33	−649.10	−981.26	−1033.53	−945.75
	Corr	0.59	0.51	0.64	0.63	0.59	0.68

the Poisson and negative binomial distributions are well documented in the related statistical literature, to the best of our knowledge, this is the first time the Poisson and negative binomial models have been explored in genomic-enabled prediction. The proposed negative binomial models were derived using the Pólya–Gamma data augmentation approach proposed by Polson et al. (2013) for count data; this approach is elegant, efficient, and leads to familiar complete conditionals on the target quantity (Windle et al. 2013). The data augmentation method is novel and consists of augmentation approaches with closed-form solutions and analytical update equations available for Gibbs sampling. One augmentation approach is concerned with the inference of the NB parameter r using compound Poisson representation, and the other approach is concerned with the inference of the regression coefficients $\boldsymbol{\beta}$ using Pólya–Gamma distribution.

Our Bayesian NB models (**Models NB** and **Pois**) for genomic-enabled prediction are different from that proposed by Polson et al. (2013), because we incorporate random effects in addition to fixed effects. For this reason, we call these models the Bayesian mixed negative binomial models. However, incorporating a random effect in the linear predictor requires that in addition to the full conditionals of ω_{ij}, $\boldsymbol{\beta}$, L_{ij}, and r, we derive the full conditionals

Table 3. Rank of the four models for the five criteria, DIC, χ^2_{cal}, L, LMPL, and Corr, resulting from the 10 random cross-validations.

Criteria	Dataset C46				Dataset C47			
	Model NB	**Model Pois**	**Model Normal**	**Model Log-normal**	**Model NB**	**Model Pois**	**Model Normal**	**Model Log-normal**
DIC	1	3	2	4	3	1	2	4
χ^2_{cal}	3	4	2	1	2	4	3	1
L	2	1	3	4	2	1	3	4
LMPL	1	4	2	3	4	1	2	3
Corr	3.5	3.5	1	2	2.5	2.5	2.5	2.5
Average rank	1.75	3	2	2.8	2.75	1.75	2.5	3

Table 4. Simulation example for two scenarios.

Scenario	Parameter	True	Model NB		Model Pois	
			Estimate	SD	Estimate	SD
1	β_0	0.66	0.64	0.11	0.64	0.14
	r	4.87	5.35	1.71	1000	–
	σ^2_u	0.55	0.61	0.14	0.60	0.17
	DIC		1476.73	66.81	1520.06	88.83
	χ^2_{cal}		300.03	11.39	478.21	48.66
	L		1044.27	103.59	904.76	78.53
	LMPL		−739.29	33.37	−770.39	46.82
	Corr		0.71	0.06	0.71	0.06
2	β_0	0.66	0.68	0.12	0.66	0.14
	r	1000	61.94	15.49	1000	–
	σ^2_u	0.55	0.56	0.14	0.59	0.13
	DIC		1380.74	54.65	1385.53	61.56
	χ^2_{cal}		300.20	17.67	320.07	24.06
	L		841.07	51.30	821.83	56.67
	LMPL		−696.31	27.85	−693.65	31.47
	Corr		0.80	0.06	0.80	0.06

In scenario 1, the data were simulated under **Model NB**, while in 2, they were simulated under **Model Pois**. For each scenario, we estimated the parameters under **Models NB and Pois** and we calculated the five criteria, DIC, χ^2_{cal}, L, LMPL, and Corr, to compare the performance of both models.

of σ^2_β, $\boldsymbol{u}$, and σ^2_u. Fortunately, even with the addition of random effects, we were able to get closed forms for conditional distributions of the parameters involved in the joint posterior distribution; this allows using Markov Chain Monte Carlo through the Gibbs sampler (Gelfand and Smith 1990). Adding the random effects to the predictor allowed us to extend the conventional BRR and GBLUP to count data (here called BRR count and GBLUP count), as was done for ordinal categorical phenotypes by Montesinos-López

et al. (2015) using the threshold model for genome-enabled prediction, which they called TGBLUP.

This extension of GBLUP to count data is very important, since it allows modeling count data in a scientific manner, without assuming that the data are normally approximated and without using transformation, which many times produces estimations and predictions outside of non-negativity, which makes no sense for count data. The BRR count is slightly different from that proposed by Polson et al. (2013), since we assumed the value of the variance of marker effects is unknown and gave this variance an inverse gamma prior distribution.

4.2. Assessing the Models' Genomic Prediction Accuracy

Our proposed models (**Models NB** and **Pois**) proved superior to **Model Normal** and the normal model with transformed data (**Model Log-normal**) for dataset C46 for criteria DIC and LMPL but not for all five criteria. **Model Pois**, however, was clearly the winner in three of the criteria for dataset C47. For dataset C46, the mean rank favored **Model NB**, whereas for data set C47, the best model was **Model Pois**. However, this finding can be attributed to the large datasets used and to the fact that a considerable proportion of the response variable had zeros in the C46 dataset. Comparing the posterior predictive mean of the number of observations in dataset C46 with the observed data when the data were fitted with **Model NB**, we did not observe the presence of an excessive number of zeros (see Fig. 4a). However, when this dataset was fitted with **Model Pois**, we observed that the number of zeros is large relative to the posterior predictive mean of the number of zeros; it seems there is evidence of zero inflation (Fig. 4b). While in dataset C47 we observed fewer zeros than expected under **Model NB** (Fig. 4c), we also see there is an excess number of twos in this dataset (Fig. 4c, d). In the presence of an excess number of zeros, an alternative modeling approach can be obtained using a zero-inflated negative binomial (or Poisson) regression. Zero-inflated models consist of a mixture where a zero-inflated structure is incorporated such that there are two classes of zeros in the count process, one coming from a point mass and the other from a non-truncated process (Boone et al. 2012). However, the Gibbs sampler proposed in this paper is not straightforward or generalizable to zero-inflated counts and should be considered for future research.

Previous studies found that the accuracy for an ordinal trait is lower than that predicted from a continuous trait of the same population (Kizilkaya et al. 2014). Our results show that the proposed models (**Models NB and Pois**) for genomic-enabled predictions of count data can also be used for modeling count data with equal mean and variance (assuming a Poisson distribution of data given the random effects) by fixing the overdispersion parameter r to a large value, such as 1000, as was done in this study. However, as mentioned in the introduction, rarely is this assumption (equal mean and variance) met in real count data. Finally, more research is needed to extend these proposed genomic-enabled prediction models to deal with so many zeros in count response variables. Further research is also required to examine the BMNB with other count data sets with a smaller sample size and a lower percentage of zeros.

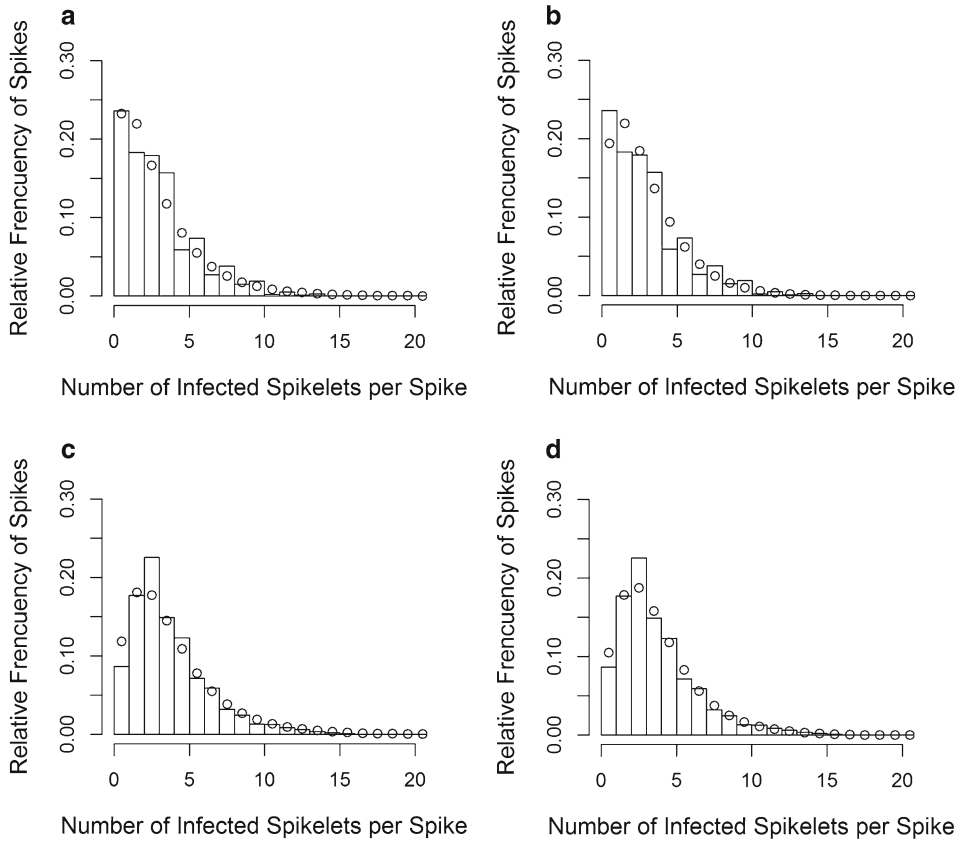

Figure 4. Histogram representation of observed counts for datasets C46 (**a**, **b**) and C47 (**c**, **d**). Superimposed as points are the posterior predictive mean for each category. The posterior predictive mean for each category was estimated under **Model NB** (**a**, **c**), and under **Model Pois** (**b**, **d**).

5. CONCLUSIONS

Genomic-enabled prediction models are useful in genomic selection for choosing candidate individuals early in life and achieving rapid genetic gains. A plethora of statistical models has been developed for genome-wise marker prediction. However, these models assume that the response variable is continuous and that its empirical distribution can be approximated by a Gaussian model. Also, the standard regression models for count data cannot deal with cases where the sample size (n) is smaller than the number of marker parameters (p). In this study, we propose a Bayesian mixed negative binomial (BMNB) regression model for count data derived using a Pólya–Gamma data augmentation for count data. We describe the conditional distributions necessary to efficiently implement a Gibbs sampler. The BMNB model (**Model NB**) together with a Poisson model (**Model Pois**), a Normal model with untransformed response (**Model Normal**), and a Normal model with transformed response using a logarithm (**Model Log-normal**) were applied to two wheat datasets (C46 and C47) using five criteria for determining the best predictive model. Based on the criteria used for assessing prediction accuracy, results indicated that the BMNB model

is a viable alternative for analyzing count data; based on one selection criterion, **Model NB** (BMNB) was the best predictive model for fitting dataset C46, whereas **Model Pois** (Poisson) was the best model for predicting data C47 based on four criteria. Results based on five criteria for assessing prediction accuracy did not determine one model to be the best based on all five criteria. Nevertheless, BMNB and the Poisson seem to be good alternative models for genomic prediction of unobserved individuals.

ACKNOWLEDGEMENTS

We very much appreciate the CIMMYT field and lab assistants and technicians who collected the data used in this study. We also thank the anonymous reviewers and the Co-Guest Editor of JABES for the time and effort they invested in correcting and improving the quality of the manuscript.

[Received March 2015. Accepted September 2015. Published Online October 2015.]

APPENDIX 1

Full conditional for β_0^*

$$
\begin{aligned}
f\left(\beta_0^*|\cdot\right) &= \left(\prod_{i=1}^{n}\prod_{j=1}^{m_i} Pr\left(Y_{ij}=y_{ij}|\boldsymbol{x}_i,r,\omega_{ij},u_i\right)\right) f\left(\beta_0^*\right)\\
&\propto \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{y_{ij}-r}{2}\left(\beta_0^*+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}+u_i\right)\right)\\
&\quad\times\exp\left(-\sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega_{ij}\left(\beta_0^*+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}+u_i\right)^2/2\right) f\left(\beta_0^*\right)\\
&\propto \exp\left(\boldsymbol{\kappa}^{\mathrm{T}}\left(\mathbf{1}_m\beta_0^*+\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{u}\right)-\frac{1}{2}\left(\mathbf{1}_m\beta_0^*+\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{u}\right)^{\mathrm{T}}\right.\\
&\left.\times\boldsymbol{D}_{\boldsymbol{\omega}}\left(\mathbf{1}_m\beta_0^*+\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{u}\right)-\frac{1}{2}\sigma_0^{-2}\left(\beta_0^*-\beta_f\right)^2\right)\\
&\propto \exp\left(-\frac{1}{2}\left[\beta_0^{*T}\left(\sigma_0^{-2}+\mathbf{1}_m^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\right)\beta_0^*-2\left(\sigma_0^{-2}\beta_f\right.\right.\right.\\
&\left.\left.\left.-\mathbf{1}_m^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{X}\boldsymbol{\beta}-\mathbf{1}_m^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Z}\boldsymbol{u}+\mathbf{1}_m^{\mathrm{T}}\boldsymbol{\kappa}\right)^{\mathrm{T}}\beta_0^*\right]\right)\\
&\propto \exp\left(-\frac{1}{2\tilde{\sigma}_0^2}\left(\beta_0^*-\tilde{\beta}_0\right)^2\right)\\
&\propto N\left(\tilde{\beta}_0,\tilde{\sigma}_0^2\right)
\end{aligned}
$$

where $\tilde{\sigma}_0^2 = \left(\sigma_0^{-2} + \mathbf{1}_m^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{\omega}} \mathbf{1}_m\right)^{-1}$, $\tilde{\beta}_0 = \tilde{\sigma}_0^2 \left(\sigma_0^{-2}\beta_f - \mathbf{1}_m^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{\omega}} \boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_m^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{\omega}} \boldsymbol{Z}\boldsymbol{u} + \mathbf{1}_m^{\mathrm{T}} \boldsymbol{\kappa}\right)$,
Full conditional for ω_{ij}

$$
\begin{aligned}
f\left(\omega_{ij}|.\right) &\propto \exp\left[-\frac{\omega_{ij}\left(\beta_0^* + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i\right)^2}{2}\right] P\left(\omega_{ij}; y_{ij} + r, 0\right) \\
&\propto \exp\left[-\frac{\omega_{ij}\left(\beta_0^* + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i\right)^2}{2}\right] P\left(\omega_{ij}; y_{ij} + r, 0\right) \\
&\propto PG\left(y_{ij} + r, \beta_0^* + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i\right)
\end{aligned}
$$

where $PG\,(b, c)$ denotes a Pólya-Gamma distribution with parameters b and c and density $P\,(\omega; b, c) = \left\{\cosh^b\left(\frac{c}{2}\right)\right\} \frac{2^{b-1}}{\Gamma(b)} \sum_{k=0}^{\infty} (-1)^n \frac{\Gamma(n+b)(2n+b)}{\Gamma(n+1)\sqrt{2\pi\omega^3}} \exp(-\frac{(2n+b)^2}{8\omega} - \frac{c^2}{2}\omega)$, where *cosh* denotes the hyperbolic cosine.
Full conditional for $\boldsymbol{\beta}$

$$
\begin{aligned}
f\left(\boldsymbol{\beta}|.\right) &= \left(\prod_{i=1}^{n}\prod_{j=1}^{m_i} Pr\left(Y_{ij} = y_{ij}|\boldsymbol{x}_i, r, \omega_{ij}, u_i\right)\right) f\left(\boldsymbol{\beta}\right) \\
&\propto \exp\Bigg(\boldsymbol{\kappa}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{\beta} - \frac{1}{2}\left(\mathbf{1}_m\beta_0^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}\right)^{\mathrm{T}} \\
&\quad \times \boldsymbol{D}_{\boldsymbol{\omega}}\left(\mathbf{1}_m\beta_0^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}\right) - \frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_v\right)^{\mathrm{T}} \boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_v\right)\Bigg) \\
&\propto \exp\Bigg(-\frac{1}{2}\Big[\boldsymbol{\beta}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{X}\right)\boldsymbol{\beta} \\
&\quad -2\left(\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2}\boldsymbol{\beta}_v - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\beta_0^* - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\kappa}\right)^{\mathrm{T}}\boldsymbol{\beta}\Big]\Bigg) \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_v\right)^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}_v^{-1}\left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_v\right)\right]\right) \\
&\propto N_p\left(\tilde{\boldsymbol{\beta}}_v, \tilde{\boldsymbol{\Sigma}}_v\right)
\end{aligned}
$$

where $\tilde{\boldsymbol{\Sigma}}_v = \left(\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2} + \boldsymbol{X}^{\mathrm{T}}\mathbf{D}_{\boldsymbol{\omega}}\boldsymbol{X}\right)^{-1}$, $\tilde{\boldsymbol{\beta}}_v = \tilde{\boldsymbol{\Sigma}}_v\Big(\boldsymbol{\Sigma}_v^{-1}\sigma_\beta^{-2}\boldsymbol{\beta}_v - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\mathbf{1}_m\beta_0^* - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\kappa}\Big)$.
Full conditional for σ_β^2

$$
\begin{aligned}
f\left(\sigma_\beta^2|.\right) &\propto \frac{1}{(\sigma_\beta^2)^{b_\beta + p/2 + 1}} \exp\left(-\frac{a_\beta + (\boldsymbol{\beta} - \beta_v)^{\mathrm{T}}\sum_v^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_v\right)}{\sigma_\beta^2}\right) \\
&\propto IG\left(a_\beta + (\boldsymbol{\beta} - \beta_v)^{\mathrm{T}}\sum_v^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_v\right)/2, b_\beta + p/2\right).
\end{aligned}
$$

Full conditional for $\boldsymbol{u}$

$$f\left(\boldsymbol{u}|.\right) \propto \exp\left(\boldsymbol{\kappa}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{u} - \frac{1}{2}\left(\mathbf{1}_m\beta_0^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}\right)^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{\omega}}\left(\mathbf{1}_m\beta_0^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}\right)\right) f\left(\boldsymbol{u}|\sigma_u^2\right)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}\left(\sigma_u^{-2}\boldsymbol{G}^{-1} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_{\omega}\boldsymbol{Z}\right)\boldsymbol{u} - 2\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{k} - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_{\omega}\mathbf{1}_m\beta_0^* - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_{\omega}\boldsymbol{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\boldsymbol{u}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{u} - \tilde{\boldsymbol{u}}\right)^{\mathrm{T}}\boldsymbol{F}^{-1}\left(\boldsymbol{u} - \tilde{\boldsymbol{u}}\right)\right\} \propto N_n\left(\tilde{\boldsymbol{u}},\ \boldsymbol{F}\right)$$

where $\boldsymbol{F} = \left(\sigma_u^{-2}\boldsymbol{G}^{-1} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_{\omega}\boldsymbol{Z}\right)^{-1}$ and $\tilde{\boldsymbol{u}} = \boldsymbol{F}\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{k} - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}_{\omega}\boldsymbol{1}_m\beta_0^* - \boldsymbol{Z}^{\mathrm{T}}D_{\omega}\boldsymbol{X}\boldsymbol{\beta}\right)$.

Full conditional for σ_u^2

If $\sigma_u^2 \sim IG\left(a_u, b_u\right)$, then

$$f\left(\sigma_u^2|.\right) \propto \frac{1}{\left(\sigma_u^2\right)^{b_u+\frac{n}{2}+1}}\exp\left(-\frac{a_u + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{u}}{\sigma_u^2}\right)$$

$$\propto IG\left(a_\sigma + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{u}/2, b_\sigma + n/2\right).$$

Full conditional for r

To make inference on r, we first place a gamma prior on it as $r \sim G\left(a_0, 1/b_0\right)$. Then we infer a latent count L for each $Y \sim NB\left(r, \pi\right)$ conditional on Y and r. Since $L \sim \text{Pois}(-r\log\left(1-\pi\right))$, by construction we can use the Gamma-Poisson conjugacy to update r. Then

$$f\left(r|.\right) \propto f\left(r\right)\prod_{i=1}^{n}\prod_{j=1}^{m_i} f\left(y_{ij}|L_{ij}\right) f\left(L_{ij}\right)$$

$$\propto r^{a_0-1}\exp\left(-rb_0\right)\prod_{i=1}^{n}\prod_{j=1}^{m_i}\left(-r\log\left(1-\pi_{ij}\right)\right)^{L_{ij}}\exp\left(r\log\left(1-\pi_{ij}\right)\right)$$

$$\propto r^{a_0+\sum_{i=1}^{n}\sum_{j=1}^{m_i}L_{ij}-1}\exp\left[-\left(b_0 - \sum_{i=1}^{n}\sum\nolimits_{j=1}^{m_i}\log\left(1-\pi_{ij}\right)\right)r\right]$$

$$\propto G\left(a_0 + \sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{m_i}L_{ij},\ \frac{1}{b_0 - \sum_{i=1}^{n}\sum_{j=1}^{m_i}\log\left(1-\pi_{ij}\right)}\right)$$

REFERENCES

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422): 669-679.

Boone, E. L., Stewart-Koster, B., & Kennard, M. J. (2012). A hierarchical zero-inflated Poisson regression model for stream fish distribution and abundance. *Environmetrics*, 23(3), 207-218.

de los Campos, G., Gianola, D., & Allison, D. B. (2010). Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat Rev Genet*, 11: 880-886. doi:10.1038/nrg2898.

de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C., & Sorensen, D. (2013a). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genetics* 9 (7) e1003608.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013b). Whole Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*, 193(2), 327-345.

Gelfand, A. E. (1996). Model determination using sampling-based methods. In: Gilks, W. R., Richardson, S., & Spiegelhalter, D. J., editors. Markov Chain Monte Carlo in practice. London: Chapman & Hall. Pp. 145-60.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398-409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian data analysis. 2. Boca Raton: Chapman & Hall.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, 473-483.

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 10: 381-391. doi:10.1038/nrg2575.

Kärkkäinen, H. P., & Sillanpää, M. J. (2012). Back to basics for Bayesian model building in genomic selection. *Genetics*, 191(3), 969-987.

Kizilkaya, K., Fernando, R. L., & Garrick, D. J. (2014). Reduction in accuracy of genomic prediction for ordered categorical data compared to continuous observations. *Genetics Selection Evolution*, 46:37 doi:10.1186/1297-9686-46-37.

Laud, P. W., & Ibrahim, J. G. (1995). Predictive Model Selection. *Journal of the Royal Statistical Society*, B 57, pp. 247-262.

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3(1), 112-115.

MacEachern, S. N., & Berliner, L. M. (1994). Subsampling the Gibbs sampler. *The American Statistician*, 48(3), 188-190.

Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., de los Campos, G., Eskridge, K. M., & Crossa, J. (2015). Threshold models for genome-enabled prediction of ordinal categorical traits in plant breeding. *G3: Genes| Genomes| Genetics*, 5(1), 1-10.

Park, T., & van Dyk, D. A. (2009). Partially collapsed Gibbs samplers: Illustrations and applications. *Journal of Computational and Graphical Statistics*, 18(2), 283-305.

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339-1349.

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink J.-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS ONE,* 7:e32253.

Quenouille, M. H. (1949). A relation between the logarithmic, Poisson, and negative binomial series. *Biometrics*, 5(2), 162-164.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet* 44: 217-220. doi:10.1038/ng.1033.

Scott, J., & Pillow, J. W. (2013). Fully Bayesian inference for neural models with negative-binomial spiking. In *Advances in neural information processing systems*, pp. 1898-1906.

Spiegelhalter, D. J., Mejor, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*, B 64, pp. 583-639.

Stroup, W. W. (2015). Rethinking the Analysis of Non-Normal Data in Plant and Soil Science. *Agronomy Journal*, 107(2): 811-827.

VanRaden, P. M. (2007). Genomic measures of relationship and inbreeding. *Interbull Bull* 37: 33-36.

——— (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414-4423.

Windle, J., Carvalho, C. M., Scott, J. G., & Sun, L. (2013). Pólya–Gamma Data Augmentation for Dynamic Models. *arXiv preprint* arXiv:1308.0774.

Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., & Simianer, H. (2014). Improving the accuracy of whole genome prediction for complex traits using the results of genome-wide association studies. *PloS One*, 9(3), e93017.

Zhou, M., & Carin, L. (2015). Negative binomial process count and mixture modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(2), 307-320.

Zhou, M., Li, L., Dunson, D., & Carin, L. (2012). Lognormal and gamma mixed negative binomial regression. In Machine Learning: Proceedings of the International Conference on Machine Learning (vol. 2012, p. 1343). NIH Public Access.